(E74-10267) AGRICULTURAL INTERPRETATION TECHNIQUE DEVELOPMENT Quarterly Progress Report (California Univ.) 15 p HC $3.00 CSCL 02D

N74-17061

Unclas
G3/13 00267

SPACE
SCIENCES
LABORATORY

E74-10267
CR-136653

AGRICULTURAL INTERPRETATION TECHNIQUE DEVELOPMENT

A report of work done by scientists at the Center for Remote Sensing Research, University of California, Berkeley, under NASA Contract No. NAS 2-7567

color

Quarterly Progress Report
30 November 1973

UNIVERSITY OF CALIFORNIA, BERKELEY

AGRICULTURAL INTERPRETATION TECHNIQUE DEVELOPMENT
(EPN NO. 382)

Principal Investigator: Robert N. Colwell
Co-Investigator: Andrew S. Benson
Contributors: Claire M. Hay,
Nancy A. Jones, Catherine E. Brown
Center for Remote Sensing Research
Space Sciences Laboratory
University of California
Berkeley, California 94720

1.0 INTRODUCTION

The emphasis of the Skylab agricultural investigations being carried out at the Center for Remote Sensing Research (CRSR) has been placed on the quantitative evaluation of the Earth Resources Experimental Package (EREP) data with respect to its usefulness in agricultural inventories. An attempt will be made to answer two basic questions: (1) Will the high quality S190A and S190B photographic data enable us to utilize manual interpretation techniques that have heretofore been considered impractical using Gemini, Apollo or ERTS data? and (2) Will the 13-channel S192 scanner data allow us to efficiently classify agricultural crops using automatic data processing techniques? The following sections of this quarterly progress report document those tasks which have accomplished during this reporting period, and those tasks which are projected for the remainder of the contract period.

2.0 WORK PERFORMED DURING THE PERIOD COVERED BY THIS REPORT

2.1 Status of Skylab Data and High Altitude Aircraft Imagery

2.1.1 Skylab 2

All photography of our study areas that was acquired from Skylab 2 (Track 63, June 3, 1973) has been received along with its supporting high altitude aircraft imagery. During the June overpass, the S190B camera was not in operation; however, high quality S190A photography was acquired.

To date, only one S192 digital tape (containing 5 seconds of data) has been received which covers the southern portion of the Salinas Valley test site (see Figure 1). Of the ten channels on this tape, four were considered unusable -- channels 1 and 13 were almost completely saturated, and a high degree of systematic noise was present in channels 6 and 11. The remaining channels -- 2, 3, 8, 9, 10 and 12 --

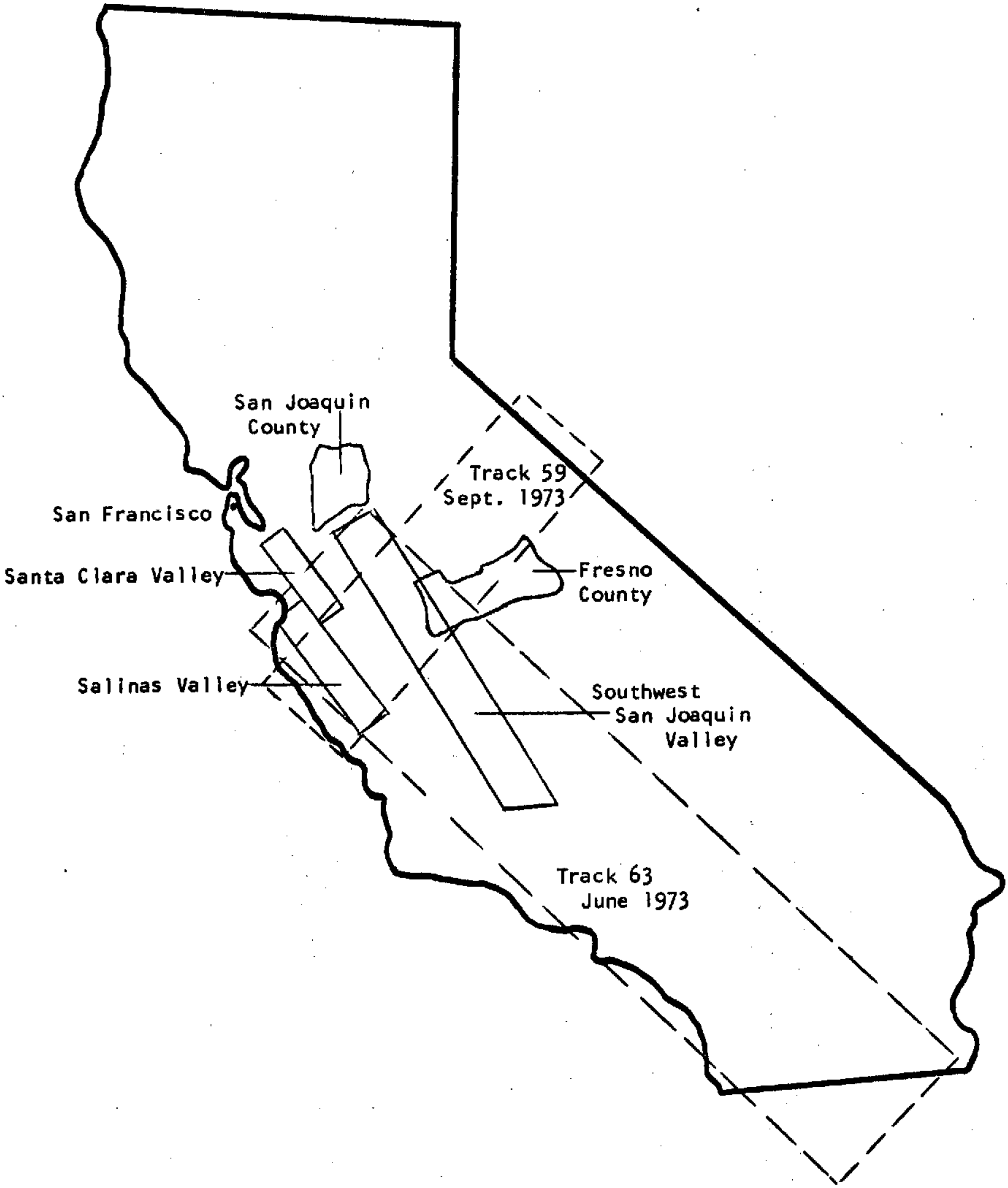


Figure 1. The above map shows the state of California and the locations of the agricultural test sites -- Santa Clara Valley, Salinas Valley, and southwestern San Joaquin Valley -- with respect to Skylab ground tracks. (Solid lines delimit test sites and dashed lines delimit S190A photographic coverage.

were usable for CALSCAN* analysis.

2.1.2 Skylab 3

September 13, 1973 was the only date during Skylab 3 on which data of the agricultural test sites were acquired. Photographic and digital data were taken on Track 59 but have not been received. Although we should receive S190A and S190B data in the near future, S192 data from Skylab 3 are not expected to be available until June, 1974.

2.2 Agricultural Land Use Stratification

The general purpose of any stratification system is to increase sampling efficiency by decreasing the variance within a set of samples. Such a decrease in variance allows for a more precise estimate to be made from those samples. With specific reference to the Skylab experiment, agricultural land use stratification from the S190A photographic images has two possible applications: (1) to facilitate automatic classification of the digitized S190A black-and-white transparencies and the S192 digital tapes, and (2) to increase the accuracy of crop inventory estimates obtained from S190A photographs using manual interpretation techniques. In the first case, it is necessary to delineate agricultural areas of homogeneous appearance in a manner that minimizes the variability in the spectral characteristics of land use types within the boundaried areas. However in the second case, the number of the resulting boundaries is often governed by the human interpreters inability to effectively use more than eight to twelve different strata types.

Two interrelated photo characteristics, color and texture, were used to differentiate between agricultural land use strata. Color differences are closely associated to crop types. For example, in the western San Joaquin Valley as seen on the June color image, light and dark buff colored areas characteristically represent the grain stubble fields, bluish gray colors are associated with the sparsely vegetated and bare soil fields, and green colors correlate with the highly vegetated fields such as sugar beets, tomatoes and melons. Texture, or the frequency with which the tone or color of the image changes per unit area, is dependent upon both the proportion of the area planted to a particular crop and the field size. Texture, therefore, generally is a good indicator of land use within an area. For example, the large black areas that appear on the June false-color infrared image are indicators of a major rice growing district which, at that time of year, is inundated with water. In addition areas with a coarse texture (field sizes of 80-160 acres) indicate a predominance of field crops, while

*CALSCAN is the CRSR version of the LARS-Purdue pattern recognition program adapted to the CDC 6600/7600 system at the University of California, Berkeley.

areas with a medium texture (field sizes of 30-80 acres) indicate a mixture of field and vegetable crops, and those with a fine texture (field sizes of 10-20 acres) indicate vineyards and pastures.

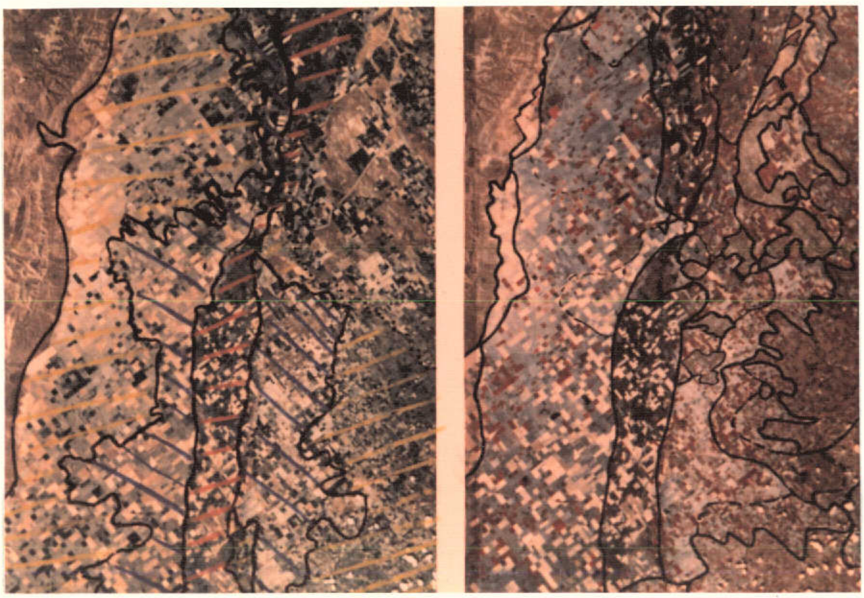
2.2.1 Stratification of the Southwestern San Joaquin Valley Test Site

Using the criteria described above, a skilled image analyst stratified a portion of the southwestern San Joaquin Valley (western half of Fresno County) into agricultural land use classes (see Figure 2). The work was done on both natural color and false-color infrared S190A imagery taken from Track 63 on June 3, 1973 (Frame #121). Strata boundaries were first delineated on enlarged 8 x 8 inch prints (scale 1:805,000). This image size-scale combination was considered optimum because it allowed the analyst to (1) discriminate between major land use areas which often are confused on larger scale imagery, and (2) delineate areas with an ink pen -- a task which would not have been possible on the original 70 mm transparency. The land use boundaries were then transferred to a 20 x 20 inch enlargement (scale 1:322,000) to facilitate their analysis and correlation with ground data and soil map data.

Of the two image types used, the false-color infrared was considered to be more useful for determining stratification units than the natural color. On the natural color image it was often difficult to differentiate between vegetated fields which appear dark green, and fallow, burned, or flooded fields which appear black. This differentiation is made quite easily on the false-color infrared image because vegetated areas appear bright red in contrast to the non-vegetated fields which are black. In addition, differences in soil moisture which occur between the different soil types are contrasted more strikingly on the color infrared than on the natural color image.

Using ground data that were collected along three transects across the study area, a chi-square analysis was performed to test the significance in terms of land use of the stratification boundaries delineated on the false-color infrared image. The results of this analysis proved that a significant interaction existed between the strata sampled and their associated land use, thus indicating that the delineations as drawn by the image analyst were indeed meaningful in terms of the objectives being sought.

The strata boundaries plotted on the false-color infrared image were compared to those appearing on local soils maps (see Figures 2a and 2b). In two cases the land use strata boundaries mapped on the Skylab imagery corresponded well with the soil boundaries. These were (1) the boundaries between the valley basin and the valley basin rim soils, and (2) the boundaries between bedrock material and soils that developed on recent alluvial fans. In the first case, the pronounced land use stratum boundary was easily mapped mainly because soil moisture differences occurring on each side of the boundary were quite



a b

Figure 2. Southwestern San Joaquin Valley as imaged on natural color (a) and false-color infrared (b) S190A imagery, June 3, 1973, frame number 121. Boundaries between major soil types have been transferred from a soil map onto photo (a) whereby the red cross hatched areas represent the valley basin soils, the blue hatching shows the valley basin rim soils, and the yellow hatching correlates with the alluvial fan soils. Unhatched areas represent soils that developed on bedrock material. The land use strata shown in photo (b) were drawn using the criteria that is described in the text, namely, image color-tone and texture. Note that land use strata shown in photo (b) are fairly well correlated with the boundaries between the valley basin and valley basin rim soils and with the boundaries between the alluvial and the bedrock soils. However, the boundary between the valley basin rim and the alluvial fan soils does not correlate with any of the agricultural land use strata boundaries. See text for further explanation.

apparent. The valley basin soils have a higher available water capacity than the valley basin rim soil and, therefore, appear darker in tone. In the second case, the correspondence between the soil and land use boundaries was due to topographic considerations which affect both land use and soil development.

2.2.2 Stratification of Salinas Valley

Using the same manual interpretation techniques as described in Section 2.2.1, the Salinas Valley test site was also stratified into major land use categories. No detailed evaluation has been made of these results as yet, but these strata will be quantitatively evaluated as an input to the CALSCAN analysis of the digitized S190A imagery.

2.3 Automatic Interpretation

The planned automatic data processing studies call for extracting agricultural information from EREP computer-compatible tapes and photographic transparencies. The processing of both types of data will give a comparison of classification results using an analog storage medium (film) versus digital storage (tape).

2.3.1 Multispectral Scanner Data

Initial experimentation with data from the S192 multispectral scanner was begun using only six channels: 2, 3, 8, 9, 10 and 12 (see Section 2.1.1). Spectral data for seven crop types, viz., immature barley, mature barley, cucumber, plowed-fallow, tomato, cotton and alfalfa, were extracted from a section of the digital tape. From a statistical analysis of these data it was determined that channels 2, 3, 8 and 12 only would be needed for the CALSCAN classifier because the data contained on channels 9 and 10 were highly correlated with those on channel 8 and therefore were considered redundant.

Of the approximately 3,500 acres (1,400 hectares) of crop land that were classified, 75 percent of the fields were identified correctly. Most of the classification errors occurred in attempting to discriminate between bare soil and cucumbers which have sparse vegetative cover. Nevertheless, these results are very encouraging when one considers that neither channel 4 (green) nor channel 6 (red) was used in the classifier. Once these two channels become available, we expect the classification accuracy to improve considerably.

2.2.2 Multispectral Camera Data

Preliminary evaluation of the digitized S190A black-and-white transparencies is continuing. The northern half of the Salinas Valley test site was scanned with a microdensitometer on all four bands of photography: infrared -- RL01 -- (.7 - .8 μm), infrared -- RL02 -- (.8 - .9 μm), red -- RL05 -- (.6 - .7 μm), and green -- RL06 -- (.5 - .6 μm). The scan interval along both the X and Y axes was .01 inches, and the aperture of the scanner was set at .001 inch diamter so that each data point represented a 1.12 acre spot size on the ground. The resulting density measurements were recorded on magnetic tape and were subsequently reformatted on the CDC 6600/7600 system so that they would be compatible with the CALSCAN program. Training fields were extracted from the reformatted tape, and the STAT sub-program* of CALSCAN was used to produce various descriptive statistics.

The mean density, standard deviation, and coefficient of variation as scanned from the four bands for tomato, asparagus, and sugar beets are shown in Table 1. (Statistics for the remaining four crop types present in the area, viz., lettuce, carrots, cauliflower, and beans, are not available at this writing due to software difficulties but will be reported in the next quarterly report.) It is quite evident from these statistics that data from one band of imagery alone will be insufficient to discriminate between these three crops. Uniform appearance, i.e., a small deviation from the mean, facilitates automatic classification. However, the crop types which display these small deviations such as tomato and asparagus on band RL02 also show a correspondingly small difference between their actual means, and thus they have virtually identical density distributions. Until data for all crop types are available it will not be possible to determine which bands of photographic data will give the most accurate discrimination of crops. However, initial reaction to these results derived from multispectral photo data is that accurate classification of these vegetable crops suing either automatic or manual techniques will be difficult, mainly because of the similar states of development of these crops at the time of photo acquisition. In June in the Salinas Valley, tomato, carrot, lettuce and cauliflower were just emerging from the soil, and had the same basic appearance as bare soil. Improvement in crop identification accuracy is expected from the SL-3 data because all crops will be fully mature. Likewise, when SL-2 (June) and SL-3 (September) data taken of the same area are combined, variations in cropping cycles will greatly facilitate discrimination between crops that have similar spectral reflectance characteristics on any single date.

3.0 WORK PLANNED FOR THE NEXT REPORTING PERIOD

3.1 Introduction

The unexpected delays experienced in receiving the SL-3 data make it difficult to accurately project the work that will be accomplished

---

*The STAT sub-program of CALSCAN performs statistical analyses on specified training fields and prints the results in various forms, as directed by the user. These data indicate which classes are separable on the basis of the training data given and also which features offer the best separation.

TABLE 1

	Tomato	Asparagus	Sugar Beets
Infrared RL01 (.7 - .8 μm)	$\overline{X}$ = 97.96	$\overline{X}$ = 109.08	$\overline{X}$ = 190.40
	S = 28.91	S = 13.65	S = 25.30
	c = .29	c = .13	c = .13
Infrared RL02 (.8 - .9 μm)	$\overline{X}$ = 129.13	$\overline{X}$ = 126.31	$\overline{X}$ = 112.86
	S = 11.97	S = 7.38	S = 41.20
	c = .09	c - .06	c - .36
Red RL05 (.6 - .7 μm)	$\overline{X}$ = 137.99	$\overline{X}$ = 150.74	$\overline{X}$ = 135.29
	S = 39.09	S = 28.58	S = 66.20
	c = .28	c = .19	c = .49
Green RL06 (.5 - .6 μm)	$\overline{X}$ = 136.22	$\overline{X}$ = 126.64	$\overline{X}$ = 138.04
	S = 40.13	S = 24.01	S = 85.30
	c = .29	c = .19	c = .62

Shown here are the mean density ($\overline{X}$), standard deviation (S), and coefficient of variation ($c = S/\overline{X}$) for three crop types as scanned from the four black-and-white EREP transparencies. These statistics will be used in part to select the optimum bands for automatic classification of crop type using the CALSCAN program.

during our next reporting period. While definitions of the tasks as outlined in the preceeding quarterly report have not changed, it is doubtful that these tasks can be completed within the time frame shown in the original work plan unless the SL-3 photographic and multispectral scanner data are made available in the very near future. Under these circumstances, and with the concurrence of our contract monitors, we will continue to devote our entire effort during the next reporting period to the analysis of SL-2 data. This work will involve intensive studies of agricultural land use stratification and crop identification using manual and automatic techniques -- but it will of necessity be limited to single date imagery taken last June.

The sections that follow below highlight several of the tasks to be performed once SL-3 data are received.

3.2 Agricultural Land Use Classification

Upon receipt of Skylab 3 photographic data, the test sites will be again stratified using the criteria described in Section 2.2. An assessment will be made as to the degree with which the resulting boundaries have shifted due to changes in cropping practices, soil moisture, etc. In addition, the strata delineated on Skylab 2 data will be put into the Center's computerized data bank for subsequent analysis using the CALSCAN program (see Section 3.3.2).

3.3 Irrigated Land Study

There presently exists an important need on the part of the Department of Water Resources of the State of California for a periodic tabulation in any given year of the statewide acreage of agricultural land receiving irrigation. A study will be performed to investigate the extent to which this tabulation can be accomplished using two dates of EREP data and the appropriate sampling designs. This study will not be able to give a valid yearly statistic of the irrigated lands within the test area due to the fact that only two dates of imagery area expected to be available and imagery taken at other dates would be required. However, it will help determine (1) the feasibility of making such inventories, (2) the expected accuracies of such inventories, (3) whether similar techniques could be used on ERTS-1 imagery, and (4) the probable cost for making such inventories on a statewide basis.

3.4 Automatic Interpretation

3.4.1 Multispectral Scanner Data

As soon as noise-corrected S192 data are provided, the optimum combination of channels needed for crop type identification will be determined. These results will then be compared with the corresponding results from the automatic classification of S190A data.

3.4.2 Photographic Data

After the initial crop type classification has been completed, the CALSCAN analysis will be rerun using the land use strata as an additional input. This work will give a quantitative measure of the usefulness of such strata boundaries for purposes of improving the efficiency (i.e., high accuracy and/or lower costs) of the automatic classification of satellite data.

3.5 Manual Crop Inventory

To date, most of the identification and inventory of agricultural crops using satellite data (i.e., ERTS-1) have been done semi-automatically by computer analysis of digital tapes with human inputs of training materials. Although such systems have achieved very accurate results in certain test areas, there is still a need to develop techniques for manual interpretation of satellite data for agricultural resource inventories. While computer-based systems may ultimately provide the most efficient method for gathering agricultural statistics of extensive areas, at the present time the human interpreter represents the most expedient way to perform an operational inventory in the United States. For many of the emerging nations of the world where both national agricultural statistics and computer systems may be non-existent, the data gathered by human intepreters can provide a valuable input to the management decisions for agricultural resources.

In any attempt to develop efficient techniques for the human interpretation of Skylab data, several factors must be considered: (1) because of the large areal coverage of the imagery, 100 percent image interpretation of the entire frame for detailed information is not practical, (2) a simple method is needed to evaluate the accuracy of the interpreter's estimates and, if necessary, to adjust these estimates, and (3) the low resolution of the imagery makes accurate acreage estimates by human interpreters impossible.

An inventory technique employing a double sampling design utilizing point data is expected to deal effectively with these constraints. For the first stage (large sample) the interpreter will determine the presence or absence of the crop of interest at each of a large number of points throughout the survey area. These data will then be used to estimate the proportion of the area that is planted to that particular crop. This proportion, when multiplied by the total area being inventoried will give an estimate of total crop acreage. The second stage in the inventory will consist of a subsample of the large sample. These subsample points will be field checked, and the correlation between ground conditions and image interpretation estimates will be used to evaluate interpreter accuracy and to calculate a ratio estimator to adjust the interpreter's estimated proportion. It must be noted that, due to the long delay between data acquisition and the receipt of EREP imagery, the second stage data will probably have to be collected from high altitude photography that was obtained coincident with the Skylab overflight.

3.6 Multistage Sampling of Agricultural Resources

In general, there are two types of basic information required by agricultural planners: (1) an estimate of the quantity of specific types of agricultural resources in each administrative unit, and (2) an inplace map of those resources. When a high correlation exists between estimates of the resources as made from spacecraft data and those based on ground observations, unique and valuable information for meeting both of these requirements can be provided in a cost-effective manner.

The culmination of our Skylab agricultural experiments will be the demonstration of agricultural survey techniques using a multistage sampling model. Through use of the discriminant analysis techniques described below, Skylab photo and scanner data will be combined with aerial photos and ground data to give an estimate of crop acreage within the areas common to Track 63 of Skylab 2 and Track 59 of Skylab 3. We have not yet chosen which crop or crops will be inventoried; this will be determined after the data from the S192 multispectral scanner have been analyzed. Under normal operational conditions, the optimum times for obtaining remote sensing data for a successful agricultural inventory are known and are carefully adhered to. However, since this is not possible under the restrictions of the EREP missions, the crops to be surveyed will be those which can best be inventoried with the available Skylab data.

The multistage model relies heavily upon the first stage in which the information extracted from the Skylab data by human interpreters and automatic classifiers provides the initial estimates of the resource. The first step of the data extraction process will be one in which human interpreters stratify all fields within the area of interest into broad land use categories and crop classes based on their appearance on the Skylab S190A ektachrome imagery. At this time political and geographic boundaries will be superimposed on the imagery to further define the areas of interest. Next, a number of fields which represent the various agricultural resources of interest will be selected from each stratum to train the discriminant analysis program. The identities of these fields will be determined from ground data and/or the interpretation of aerial photos. The number of training fields required for each crop class will be dependent upon the variability of the spectral signatures of the crops present. This variability is caused by such factors as different cropping practices, local soil differences, and genetic variations within each particular crop type. For example, a crop such as alfalfa which may be in several stages of maturity throughout the survey area at the time of image acquisition may require five or more fields per stratum for adequate training, whereas only one training field per stratum may be needed for a less complex crop type such as corn. After the fields have been chosen, they will be located on, and extracted from, the spacecraft imagery. The multispectral data from the training fields will be run through the discriminant analysis program to obtain a point-by-point classification of the entire area by strata. This procedure will provide an initial estimate of the acreage of the vegetation classes.

In the second stage of the model, the results of the discriminant analysis will be sampled to determine their relationship to ground estimates of the resource. Sampling units (SU) will be defined by dividing the entire area into rectangular blocks. The size and shape of these blocks will be determined by (1) the information requirements, (2) the variability that occurs in the SU estimates as the block size changes, (3) the costs of further subsampling, and (4) the resolution of the scanner imagery.

Probability sampling is expected to provide the most efficient sampling design that can be applied in the second stage of the model. Probability sampling is a special case of the mean of the ratios estimation where samples are allocated proportional to the expected variance of the X_i estimate. For this model, the total value of the i^{th} SU, denoted by X_i, is evaluated by

$$X_i = \sum_{m=1}^{M} \sum_{j=1}^{J} I_m V_j$$

where $\left| \begin{array}{l} I_m = 1 \text{ if } C_m = j \\ I_m = 0 \text{ otherwise} \end{array} \right.$

C_m = crop class for the n^{th} "pixel" (picture element) of the SU, as determined by the discriminant analysis,

M = the number of "pixels" per SU,

V_j = the crop class being evaluated

J = the number of crop classes.

The value or weight (v_j) is assigned to rank the various crops or vegetation types based on their relative importance to the survey. In an agricultural inventory where total dollar value is the objective, the v_j's are the average dollar values per "pixel" of the various crops (j). If the making of an inventory of a single crop is the objective, the value of the crop of interest is 1 and all other v_j's are set to zero. In many cases this weighting factor is primarily affected by the marketing conditions of each crop, and is highest for those crops for which acreage estimate errors are most important.

The variance of the population is estimated by

$$S^2 = \frac{1}{N-1} \sum (X_i - \bar{X})^2$$

The number (n) of SU's to be selected for photo and ground measurement when no remote sensing information is available is determined by

$$n = \frac{Nt^2S^2}{N(AE)^2 + t^2S^2}$$

where AE = the allowable error, in units of value

t is a value obtained from "students t" tables and

S^2 is as defined previously.

The n points are then selected from the list of SU's proportional to their estimated value.

The selected SU's are then carefully transferred to the corresponding high flight photography where precise field size measurements are made for use later in adjusting the acreage estimates obtained from the classifier.

From high flight images, low altitudes images, ground identification and historical data, the "correct" classification for each field in the SU is determined, down to crop type and maturity.

The total value for the area ($\hat{T}$) is estimated using the probability of selection (P_i) and the photo/ground estimate of SU value (Y_i) by means of the relation,

$$\hat{T} = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i}{P_i}$$

where $$P_i = \frac{x_i}{\sum_{i=1}^{N} x_i}$$

The variance of the estimate for $\hat{T}$ is

$$s_{\hat{T}}^2 = \mathrm{Var}(\hat{T}) = \frac{1}{n} \sum_{i=1}^{N} P_i \left(\frac{Y_i}{P_i} - \hat{T}\right)^2$$

If the photo/ground estimate (Y_i) were to be perfectly proportional to the remote sensing estimate (X_i), only one ground sample would be needed to determine the proportionality constant. More realistically however, the number of ground samples (n) for future surveys is estimated by:

$$n = \frac{Nt^2 s_{\hat{T}}^2}{N(AE)^2 + t^2 s_{\hat{T}}^2}$$

This probability sampling model is appropriate when a single parameter such as "acreage of a single crop", "value of all the crops present", or "demand for irrigation water" is desired. Such a model can be replaced by a regression sampling model if estimates on a crop-by-crop basis are required; however, the regression sampling model will only meet the allowable error criterion for the total value of all crops by strata.